



Publishing Uncertainty on the Semantic Web: Blurring the LOD Bubbles

Ahmed El Amine Djebri, Andrea G B Tettamanzi, Fabien Gandon

► To cite this version:

Ahmed El Amine Djebri, Andrea G B Tettamanzi, Fabien Gandon. Publishing Uncertainty on the Semantic Web: Blurring the LOD Bubbles. ICCS 2019 - 24th International Conference on Conceptual Structures, Jul 2019, Marburg, Germany. pp.42-56, 10.1007/978-3-030-23182-8_4. hal-02167174

HAL Id: hal-02167174

<https://inria.hal.science/hal-02167174>

Submitted on 27 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Publishing Uncertainty on the Semantic Web: Blurring the LOD bubbles

Ahmed El Amine Djebri, Andrea G.B. Tettamanzi, and Fabien Gandon

Université Côte d’Azur, Inria, CNRS, I3S, France
djebri.emp@gmail.com
andrea.tettamanzi@univ-cotedazur.fr
fabien.gandon@inria.fr

Abstract. The open nature of the Web exposes it to the many imperfections of our world. As a result, before we can use knowledge obtained from the Web, we need to represent that fuzzy, vague, ambiguous and uncertain information. Current standards of the Semantic Web and Linked Data do not support such a representation in a formal way and independently of any theory. We present a new vocabulary and a framework to capture and handle uncertainty in the Semantic Web. First, we define a vocabulary for uncertainty and explain how it allows the publishing of uncertainty information relying on different theories. In addition, we introduce an extension to represent and exchange calculations involved in the evaluation of uncertainty. Then we show how this model and its operational definitions support querying a data source containing different levels of uncertainty metadata. Finally, we discuss the perspectives with a view on supporting reasoning over uncertain linked data.

Keywords: Uncertainty · Linked Data · Semantic Web.

1 Introduction

Many applications consume data from the World Wide Web either as main data to perform some analysis or as additional data to augment internal databases. However, the Web is an open world where everyone can state any information, with the possibility of altering existing assertions and publishing incorrect, ambiguous, misleading or uncertain information. In addition, users can be biased against controversial sources that might afford correct and valid information, and blindly trust some other sources. The lack of objectivity and the deficiencies of the Web are projected onto the Semantic Web [1] and the Linked Data [2], as the latter is mostly populated by harvesting existing Web resources. For instance, a huge amount of data in *Freebase* [3] was imported automatically and independently of the approval of the community. Before these external sources may be used in data mining or other data processing, their reliability must be assessed.

Uncertainty might differ from one use case to another in terms of nature and requirements. It can take the form of inconsistencies, incompleteness, ambiguity,

vagueness, etc. The use of such uncertain data requires a specific representation which the current standards of the Semantic Web of Linked data are not able to fulfill without extensions. Although the Semantic Web technologies have been adopted in many applications over the last two decades—recommender systems [4], news aggregators [5], and even the legal domain [6]—operating with uncertain data in such domains raises multiple questions. The main question we address in this paper is: *Can uncertainty be formalized on top of the standards of the Semantic Web to be published on the Web?*

To answer this question, we present a new vocabulary for the representation of uncertainty in the Semantic Web. The **meta-Uncertainty** vocabulary enables the representation of custom uncertainty approaches using standard RDF and the annotation of sentences and worlds using uncertainty values. It also allows the publishing of uncertainty calculi, enabling the reusability of uncertainty approaches. Afterward, we present an original framework enabling multi-level uncertainty mapping to sentences. The latter is mapped to a combination of their own uncertainty information and that of their context.

Section 2 describes the related work and Section 3 focuses on motivating scenario and the formalization of uncertainty in the Semantic Web. Section 4 addresses querying and mapping metadata about uncertainty. Section 5 is a discussion of the current state of our work before we conclude with perspectives.

2 Related Work

The study of deficiencies and uncertainty of information issued from the Web has been the object of a non-negligible part of the research work of the Semantic Web community. One of the main challenges [7] of interconnected uncertain data sources is the representation of uncertainty itself. The standardization is crucial for interoperable applications relying on the uncertain Web in which some sources have already an application-specific representation for uncertainty.

Uncertainty in the literature can be represented in many theories, each having a formal background, a set of measures and a defined calculus. Dubois et al. [8] present an overview of uncertainty representations in Artificial Intelligence and Decision Theory and discuss multiple theories, such as probability theory, the theory of evidence of Dempster-Shafer and possibility theory. Klir et al. [9] observe that dealing with uncertainty consists of four distinct levels: representation, calculus, measurement, and development of methodological aspects of the theory. Our focus in this paper is on the first two levels.

To adapt the previous uncertainty representations to fit within the Semantic Web Standards, the W3C Incubator Group URW3-XG proposed an ontology to formalize data annotation with uncertainty information [10]. The work focuses on the type and the nature of uncertainty models. However, the ontology proposed a limited set of theories for which metrics for comparison or evaluation are not offered. That does not allow evaluating uncertainty information nor comparing two uncertain statements to tell which one is more reliable. Another work presents a vocabulary for imprecise values [11]. The aim was to use the

new definition to solve data fusion problems by extending RDF semantics [12] with source and interval information. This vocabulary offers to handle fuzzy or inconsistent data and is useful in time-based reasoning. Yet this vocabulary does not provide a formal definition of uncertainty.

Another part of the literature focuses on extending the current standards of the Semantic Web to fit in the existing uncertainty theories. Some authors presented probability as a *threshold* [13] or a *confidence score* [14] linked to a graph pattern. Other authors extended the OWL standard by linking a *degree of truth* with statements to enable representing Bayesian Networks [15]. More extensions propose to handle fuzzy values [16] and possibility theory [17] each using a specific set of measures depending on the theory. In general, each of these previous works focuses only on one uncertainty theory and defines a set of measurements (features) to rely on while ranking, reasoning, or aggregating uncertain data. However, limiting the representation of uncertainty to one approach per application requires implementing as many extensions as existing approaches. In that case, the problem of sharing and reusing data between the different systems persists.

To the best of our knowledge, there is no work that offers to publish uncertainty theories alongside their features and calculus. Interesting work from Dividino et al. [18] introduced the meta-knowledge mapping to statements using a non-standard representation (Quintuples) and two separated graphs (one for knowledge, another for meta-knowledge). Their framework enables the system administrator to define meta-knowledge properties and, for each property, one should define the intended semantics and the knowledge dimension inside the application in a non-standard format.

From another angle, the use of uncertain information invokes the question about context, since the reliability of information may differ from one context to another. A good reference about the importance of a context while reasoning about data validity can be found in [19]. We are interested in the case where data stated in a context can inherit its metadata.

Currently, the closest standard representation of contexts in Semantic Web is Named Graphs [20]. Data can be encapsulated inside a named graph, referenced by a URI, which can be annotated by metadata. Such representation also enables expressing and reasoning with different viewpoints without creating conflicts or inconsistencies. However, named graphs have no defined semantics in RDF other than being named containers inside the default graph. Hence, annotating a named graph does not reflect on the contained data as there is no semantic mapping between the data and their named graph. As mentioned in [21], named graphs in RDF datasets are irrelevant to determining the truth of a dataset as the latter depends on the RDF interpretation of the default graph. The document also indicates the necessity to extend RDFS [22] to support ternary relations. A simple alternative to named graphs is reification, allowing meta-statements about statements. Nevertheless, the reified statement does not necessarily assert the original one and the same problem about the propagation of the metadata from the associated context is present. A proposal in [23] approves the use of

named graphs as contexts and argues that the former are already defined by a URI. In addition, the proposal extends the semantics of RDF to use inheritance in context building. Hence, the sub-contexts are more specific than the super-contexts. Nevertheless, in order for this to work at a triple level, each and every triple should have its own context with personalized metadata. Moreover, this work considers only overriding general metadata with more specific ones, hence the logic of uncertainty theory is not applied since a triple may have different aggregation and conflict resolution techniques for each uncertainty features.

3 Formalizing and Publishing Uncertainty on the Semantic Web

The first step in publishing uncertainty is to support its representation in a way that can be queried and processed. The formalism and the calculation allowed by this representation must provide results together with uncertainty meta-data that are helpful to the user in making decisions. In particular, the model we target must account for the use of different theories to represent uncertainty – for instance across different sources – and the calculus required within and between them – for instance when joining or aggregating data.

In this section we define the *mUnc*¹ vocabulary, formalized in OWL. Then we introduce an extension based on the *LDScript* function definition language [24] to define the calculus attached to a given uncertainty theory. Finally, we provide the first validation of the expressivity of our model w.r.t. the state of the art by showing how it can capture very different types of uncertainty theories.

3.1 *mUnc*: a Vocabulary for Uncertainty Theories

mUnc (for *meta-Uncertainty*) is an OWL ontology for representing uncertainty metadata. It enables publishing uncertainty information based on uncertainty theories. Figure 1 gives an overview of the core concepts and properties of *mUnc* including: sentences, contexts (worlds) and uncertainty metadata (theories, features, calculi).

We have taken the definition of *sentence* and *world* proposed by the URW3-XG. A *sentence* is an expression evaluating a truth value, while the *world* represents the context in which a *sentence* is stated. Both sentences and worlds, can be annotated with *meta* information. For instance, the sentence `ex:S1` representing the triple `<ex:StefanoTacconi, dbo:height, 188>` referring to the height of the football player is stated in the context of the French language chapter of DBpedia [25], assuming that the latter is consistent [26]. Uncertainty information is considered a specialization of the general concept of metadata. This simplifies the future extensions for other types of metadata such as provenance or trust. For the same reason, we do not include the concept of *Agent*, as it can be included using other vocabularies like W3C *PROV*² Ontology.

¹ <http://ns.inria.fr/munc/>

² <http://www.w3.org/TR/prov-o/>

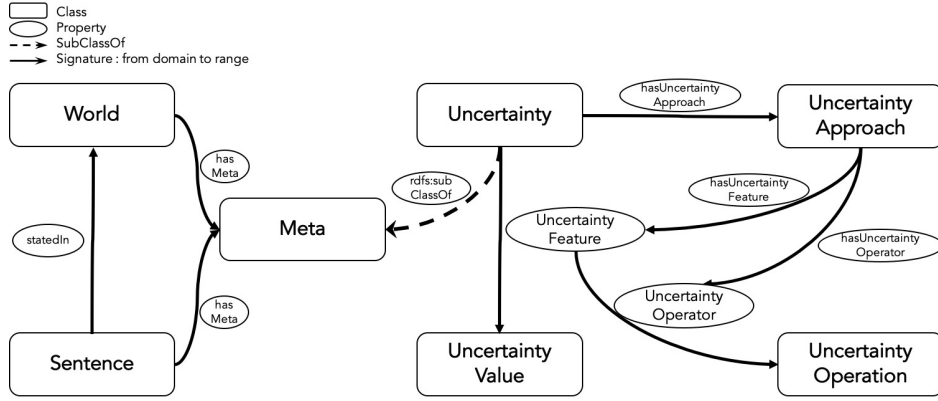


Fig. 1. Overview of the *mUnc* ontology and its core concepts

An uncertainty theory (*Uncertainty Approach*) is linked to a set of *features* and *operators*. The *features* are the metrics on which uncertainty theory is based to indicate the degree of truth, credibility, or likelihood of a sentence. Each *feature* links a *value* to the *uncertainty* information. The *operators* represent the logic to apply to the previous *values*, while the *operations* are the implemented calculus for such logic. Other concepts in URW3-XG like the type or the derivation of uncertainty can be represented as features of an uncertainty approach.

To illustrate the previous definitions, we can annotate the previous sentence using probability theory. The latter can be represented using only one feature: the probability value. We choose three logical operators to include with the definition: and, or, not. Listing 1.1 shows how to assert that a sentence `ex:S1` is true with a probability of 0.7.

```
ex:Probability a munc:UncertaintyApproach;
  munc:hasUncertaintyFeature ex:probabilityValue;
  munc:hasUncertaintyOperator ex:and, ex:or, ex:not.

ex:probabilityValue a munc:uncertaintyFeature;
  rdfs:range xsd:decimal.

ex:S1 munc:hasMeta [a munc:Uncertainty;
  munc:hasUncertaintyApproach ex:Probability;
  ex:probabilityValue 0.7].
```

Listing 1.1. Representing the Probabilistic approach using *mUnc*

3.2 Attaching Uncertainty Calculi to Theories

Semantic Web ontology languages focus on classification ontological knowledge and do not support the provision of procedural attachments or functions inside

ontologies. Our model allows linking the features of uncertainty approaches to their proper calculi (arithmetic, logical, comparison, etc). To represent the calculi, we rely on the *LDScript* function definition language [24]. This language is an ideal candidate as it is based on the SPARQL filter language and allows defining and using extension functions directly inside the Semantic Web framework. Using *LDScript*, we can define functions having a URI as a name and one or several arguments that are variables in SPARQL syntax. This enables defining uncertainty operations and link them to features using operators. To continue with the previous example, considering the fact that the sentence `ex:S1` is true with a probability of 0.7, and is stated in a context `ex:C1` in which all facts are considered true with a probability of 0.9, the conjunction of the two probabilities, assumed to be independent, is calculated using the formula

$$P(A \wedge B) = P(A) \times P(B). \quad (1)$$

The conjunction of the two previous values can be presented to the user using the function `ex:MultiplyIndependentProbability` defined using *LDScript* as shown in Listing 1.2.

```
function ex:MultiplyIndependentProbability(?pA, ?pB){
  ?pA * ?pB
}
```

Listing 1.2. Conjunction of Independent Probabilities using *LDScript*

Therefore, binding the function `ex:MultiplyIndependentProbability(0.7, 0.9)` during a SPARQL query execution will return 0.63. The former definition of the probabilistic approach using *mUnc* can be enriched by linking the URI of the function to the declared feature:

```
ex:ProbabilityValue ex:and ex:MultiplyIndependentProbability.
```

3.3 Validating the Expressivity on State of the Art Theories

As a validation of our approach, we show in this section how it supports the use of different theories to express uncertainty. A first example with the probability theory was presented in the previous section.

The enumeration of the features linked to an uncertainty theory and the comprehension of their calculus (operations) enable *mUnc* to represent existing uncertainty theories that are defined in the literature [9, 8, 27].

We can annotate the previous example using a possibilistic approach, with *validity* and *completeness* as features, as proposed in the work by da Costa Pereira et al. [28]. In Listing 1.3 we specify also the types for values linked to the proposed features:

```
ex:Possibility a munc:UncertaintyApproach;
  munc:hasUncertaintyFeatures ex:validity, ex:completeness;
  munc:hasUncertaintyOperators ex:and, ex:or, ex:not.
ex:validity, ex:completeness a munc:uncertaintyFeature;
```

```
rdfs:range xsd:decimal.
```

Listing 1.3. Possibility theory using *mUnc*

The operations for such features will be different as the conjunction of two *validity* values is their maximum. The latter can be evaluated by the function `ex:Max` defined in the listing 1.4. This function can be defined once and linked to any uncertainty approach.

```
function ex:Max(?v1, ?v2){
  if(?v1 >= ?v2) { ?v1 }
  else { ?v2 }
}
```

Listing 1.4. Conjunction of *validity* measures using *LDScript*

In this example we consider a sentence `ex:S2` with no uncertainty information, and another sentence `ex:S3` imported from another context `ex:C2` having uncertainty information. The latter can be annotated with metadata about the provenance, which can be provided using *PROV* vocabulary. This enables the traceability of uncertainty information.

```
ex:S2, ex:S3 munc:statedIn ex:C1.

ex:S3 munc:hasMeta [a munc:Uncertainty;
  munc:hasUncertaintyApproach ex:Possibility;
  ex:validity 1;
  ex:completeness 0.3
  prov:wasDerivedFrom ex:C2].

ex:C1 munc:hasMeta [a munc:Uncertainty;
  munc:hasUncertaintyApproach ex:Possibility;
  ex:validity 0.8;
  ex:completeness 0.6].
```

Listing 1.5. Context Annotation

Besides the existing uncertainty theories, *mUnc* can be used to define custom approaches to handle measures (datatypes) that cannot fit in the RDF standard. For example, we can define a custom approach to deal with fuzzy values using lower and upper bounds as uncertainty features, and the several functions about fuzzy values as uncertainty operations.

In the next section, we will discuss how to link between uncertainty information and sentences during the query processing.

4 Querying for Uncertain Linked Data

The presentation of uncertain linked data to the end user requires more information to be provided during the query processing, enabling him or her to rank and choose among the given answers. Since each sentence and context has its

own uncertainty information, if a sentence is stated in a context, uncertainty information from both should be combined to form a result. For instance, if a sentence S in a context A has no uncertainty information, it inherits uncertainty information from A (noted as \mathcal{U}_A). Otherwise, S will be annotated with a combination of both \mathcal{U}_{S_A} and \mathcal{U}_A .

4.1 Semantics of RDF with mUnc

Our work focuses on querying n uncertain data sources s_1, s_2, \dots, s_n , each possibly containing several named graphs $G_{ij}, i \in \{1, \dots, n\}$ representing each a set of coherent information. This means that each named graph contains a set of triples that do not lead to a contradictory reasoning. For example, if a named graph contains a triple `(dbo:height, rdf:type, owl:FunctionalProperty)`, it cannot have both triples `(ex:Person1, dbo:height, 188)` and `(ex:Person1, dbo:height, 195)`. We still can declare both triples in different named graphs. We recall here the definition of RDF Dataset and perform the link with the set of contexts.

Definition 1. (*RDF dataset*) An RDF dataset of a source s_i is a collection of RDF graphs, containing one default graph G_i and a set of named graphs, each consisting of a pair (u_j, G_{ij}) where u_j is the IRI of the graph G_{ij} . The set of named graphs can be the empty set.

As cited in [21], named graphs are suitable for context representation as they allow encapsulating a set of triples in a graph and annotate the latter with metadata. Also, each named graph can represent a vision or an opinion over the reality represented in the source. A sentence can be cited in multiple named graphs but with different uncertainty information. For example, two websites can state that tomorrow it will rain. The two websites may not be sure about that information at different levels, so they annotate the fact with different uncertainty information. Each of the previous websites can be encapsulated in a named graph representing a *context*.

Definition 2. (*Context*) A context $C_{ij}, j \geq 0$ is a named graph $(u_i, G_{ij}), j \geq 0$ in the RDF Dataset of a source s_i .

Each context can be annotated with a set of uncertainty information triples defined as follows.

Definition 3. (*Context Uncertainty*) A context uncertainty $\mathcal{U}_{C_{ij}}$ is a set of pairs $(\text{UncertaintyFeature}, \text{UncertaintyValue})$ representing the uncertainty information about the context $C_{ij}, j \geq 0$ in a datasource s_i .

Triples in the default graph of the source s_i may present a context on their own, and they are moved to a named graph G_{i0} representing a separate context C_{i0} . The set of pairs $(\mathcal{U}_{C_{ij}}, C_{ij})$ represents the *contextual dataset* (noted as $CDS(s_i)$) of datasource s_i .

Definition 4. (*Contextual Dataset*) Given a datasource s_i and a set of Contexts $C_{ij}, j \geq 0$, each annotated with a set of metadata triples $\mathcal{U}_{C_{ij}}$, a contextual dataset $CDS(s_i)$ of a datasource s_i is a set where C_i is the default context encapsulating metadata about other contexts, C_{i0} is the context encapsulating triples which was stored in the default graph G_i of datasource s_i .

$$CDS(s_i) = \{(\mathcal{U}_{C_{ij}}, C_{ij}) \mid j \geq 0\} \quad (2)$$

Figure 2 illustrates the definition of contextual datasets.

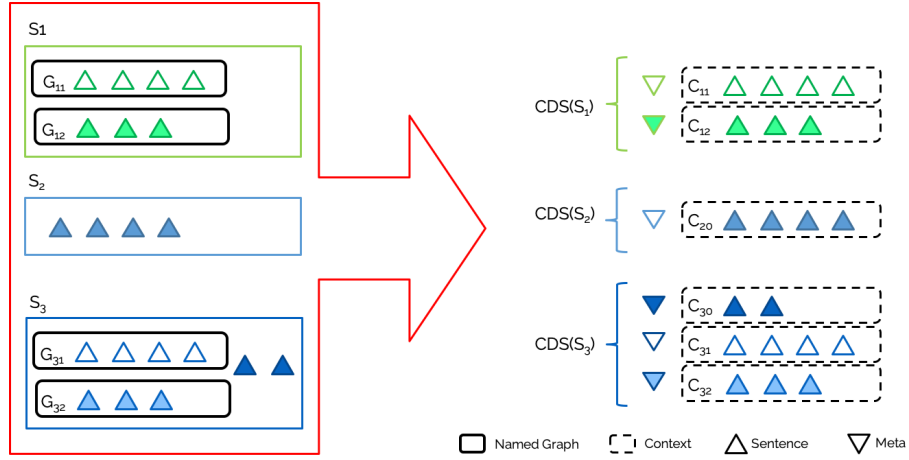


Fig. 2. Example of context encapsulation

Similar to context uncertainty, a sentence can have its own uncertainty information defined as follow.

Definition 5. (*Sentence Uncertainty*) A sentence uncertainty $\mathcal{U}_{S_{C_{ij}}}$ is a set of pairs (*UncertaintyFeature*, *UncertaintyValue*) representing the uncertainty information about the sentence S in a context C_{ij} .

mUnc does not provide any extension of RDF Semantics. Instead, we rely on the SPARQL query language to provide a mapping between sentences and the uncertainty information presented to the user. Moreover, we consider *mUnc* as an approach to provide definitions of known and custom uncertainty theories, for which we do not provide any specific semantics. The possibility of defining a calculus alongside with the ontology is an alternative to generalize and to make reuse of the shared rules between uncertainty theories such as maximizing or minimizing a feature.

This process does not require a specific RDF representation (triple, quad, etc.) on data sources. It grants the possibility to unify data represented with different syntaxes (triples, quads, etc) and from several sources. For this, we extend

the *Corese* semantic web engine [29] to enable loading and creating contexts in a transparent way. The process is executed while loading RDF data into *Corese*.

4.2 Query Resolution

In a semantic Web supporting the representation of uncertainty metadata, query results should include information about their uncertainty. The latter should be dependent on both sentences and their context.

Mapping each sentence S with its uncertainty information requires defining a *metadata-mapping mode* (see table 1) to link between the metadata annotating the sentence and the one annotating the context in which the sentence is stated.

Definition 6. (*Meta-Mapping Mode*) Given two sets $A = \{(x, y) \mid x \in F_1\}$, $B = \{(w, z) \mid w \in F_2\}$. A meta-mapping mode is the process linking A and B to a new set C where $C = \{(f, v) \mid (f, v_1) \in A, (f, v_2) \in B, v = v_1 \oplus v_2\}$.

The meta-mapping modes are basically about choosing a set of uncertainty features, and what uncertainty value to associate to each feature based on their associated values in each of the initial sets. In the first two modes, only uncertainty information $\mathcal{U}_{C_{ij}}$ linked to the context C_{ij} is considered with a specification of the reading (distributive or collective) [18]. The third mode considers uncertainty information from the lowest level of granularity only, while the fourth mode enables inheriting context metadata but overrides the values for existing features in uncertainty information linked to the sentence.

| Considered level of granularity | Mode | Explanation |
|---------------------------------|----------------------|--|
| Context | Inheritance | $\hat{\mathcal{U}}_{S_{C_{ij}}} = \mathcal{U}_{C_{ij}}$ |
| Context | Distributive reading | $\mathcal{U}_{C_{ij}} = Comb(\{V_1, V_2, \dots, V_{ C_{ij} }\}), (F, V) \in \hat{\mathcal{U}}_{S_{C_{ij}}}$ |
| Lowest granularity only | Override-All | $\hat{\mathcal{U}}_{S_{C_{ij}}} = \begin{cases} \mathcal{U}_{S_{C_{ij}}} & \text{if } \mathcal{U}_{S_{C_{ij}}} \neq \emptyset \\ \mathcal{U}_{C_{ij}} & \text{else} \end{cases}$ |
| Lowest granularity first | Override if Exists | $\hat{\mathcal{U}}_{S_{C_{ij}}} = \mathcal{U}_{S_{C_{ij}}} \cup \{(F, V) \mid (F, V) \in \mathcal{U}_{C_{ij}}, F \notin \mathcal{U}_{S_{C_{ij}}}\}$ |
| All | Combine | $\hat{\mathcal{U}}_{S_{C_{ij}}} = \{(F, V) \mid (F, V) \in \mathcal{U}_{C_{ij}}, \nexists V' (F, V') \in \mathcal{U}_{S_{C_{ij}}}\} \cup \{(F, V) \mid (F, V) \in \mathcal{U}_{S_{C_{ij}}}, \nexists V'' (F, V'') \in \mathcal{U}_{C_{ij}}\} \cup \{(F, V) \mid \exists V_{C_{ij}}, V_{S_{C_{ij}}} (F, V_{C_{ij}}} \in \mathcal{U}_{C_{ij}}, (F, V_{S_{C_{ij}}}) \in \mathcal{U}_{S_{C_{ij}}}, V = eval(Calculus(F), V_{C_{ij}}, V_{S_{C_{ij}}}))\}$ |

Table 1. Metadata-mapping modes

The set of uncertainty information linked to a sentence S regarding its context is denoted as *Universal Uncertain Information Set* or $\hat{\mathcal{U}}_{S_{C_{ij}}}$ and defined as follows.

Definition 7. (*Universal Uncertain Information Set*) A universal uncertain information set $\hat{\mathcal{U}}_{S_{C_{ij}}}$ of a sentence S in a context C_{ij} of the datasource s_i is a set of (*UncertaintyFeature*, *UncertaintyValue*) pairs issued by combining the sets $\mathcal{U}_{C_{ij}}, \mathcal{U}_{S_{C_{ij}}}$ using a meta-mapping mode.

The mapping between a sentence S and its universal uncertainty information set $\hat{\mathcal{U}}_{S_{C_{ij}}}$ is a two steps process:

- Mapping both sentences and contexts to their uncertainty information. We denote $\mathcal{U}_{S_{C_{ij}}}$ the uncertainty information about the sentence S cited in the context C_{ij} and $\mathcal{U}_{C_{ij}}$ the uncertainty information about the context C_{ij} .
- $\mathcal{U}_{S_{C_{ij}}}$ is combined with $\mathcal{U}_{C_{ij}}$ using uncertainty operations linked to each feature, in order to evaluate its corresponding value in $\hat{\mathcal{U}}_{S_{C_{ij}}}$. In this step we apply the *metaList* algorithm (see Algorithm 1) translated from the formula in the fifth meta mapping-mode (see table 1): $\hat{\mathcal{U}}_{S_{C_{ij}}} = \text{metaList}(\mathcal{U}_{S_{C_{ij}}}, \mathcal{U}_{C_{ij}})$.

Algorithm 1 *metaList*: Universal Uncertainty Information Set of a sentence S in a context C_{ij} of a datasource s_i

```

1: procedure METALIST( $\mathcal{U}_{S_{C_{ij}}}, \mathcal{U}_{C_{ij}}$ )
2:   uncFeatures( $x$ ): list of all uncertainty features contained in the set  $x$ 
3:   uncValue( $f, x$ ): the value linked to the feature  $f$  in the set  $x$ 
4:   Operation( $f$ ): the uncertainty operation linked to the feature  $f$ 
5:   eval( $o, v1, v2$ ): execute the operation  $o$  passing the parameters  $v1, v2$ 
6:    $\hat{\mathcal{U}}_{S_{C_{ij}}} \leftarrow \emptyset$ 
7:   if  $\mathcal{U}_{S_{C_{ij}}} \neq \emptyset$  then
8:     for all  $f \in \text{uncFeatures}(\mathcal{U}_{S_{C_{ij}}}) \cap \text{uncFeatures}(\mathcal{U}_{C_{ij}})$  do
9:        $v \leftarrow \text{eval}(\text{Operation}(f), \text{uncValue}(f, \mathcal{U}_{S_{C_{ij}}}), \text{uncValue}(f, \mathcal{U}_{C_{ij}}))$ 
10:       $\hat{\mathcal{U}}_{S_{C_{ij}}} \leftarrow \hat{\mathcal{U}}_{S_{C_{ij}}} \cup \{(f, v)\}$ 
11:    end for
12:    for all  $f \in \text{uncFeatures}(\mathcal{U}_{C_{ij}}) \setminus (\text{uncFeatures}(\mathcal{U}_{S_{C_{ij}}}) \cap \text{uncFeatures}(\mathcal{U}_{C_{ij}}))$ 
13:       $v \leftarrow \text{uncValue}(f, \mathcal{U}_{C_{ij}})$ 
14:       $\hat{\mathcal{U}}_{S_{C_{ij}}} \leftarrow \hat{\mathcal{U}}_{S_{C_{ij}}} \cup \{(f, v)\}$ 
15:    end for
16:  else
17:     $\hat{\mathcal{U}}_{S_{C_{ij}}} \leftarrow \mathcal{U}_{C_{ij}}$ 
18:  end if
19:  return  $\hat{\mathcal{U}}_{S_{C_{ij}}}$   $\triangleright$  the set of universal uncertainty information of the sentence  $S$ 
20: end procedure

```

In our case, *Uncertainty Operations* are stored as Linked Functions. This feature in *Corese* [29] enables storing *LDScript* [24] functions in external SPARQL query files on the web to be called at the moment of query execution. The former

feature permits publishing and executing the calculi of uncertainty approaches. Additionally, this approach may be extended to allow the capitalization of existing software libraries from other programming languages like C++ or Java.

5 Discussion

Compared to the previously published approaches, our work proposes an ontology to represent and formalize uncertainty, which covers also calculus. We rely on RDF standards to present existing uncertainty approaches in a machine-readable and reusable format. Our vocabulary also enables users to define their own uncertainty approaches and publish their definitions alongside with their uncertain data, allowing the reusability of the information in other Semantic Web applications.

The current representation of contexts using named graphs is limited because it requires the materialization of the graphs. Also, some sentences might be redundant in different contexts. A possible alternative would be the enrichment of RDF semantics to use `munc:statedIn` as an equivalent to storing sentences in named graphs. This gives more flexibility to the process and allows defining new methods and terms allowing context-overlapping and context-selective querying.

mUnc also enables representing uncertainty about uncertainty information, by considering the latter as sentences with uncertainty. Yet the combination of uncertainty information said about uncertainty sentences with the information provided by the sentences themselves is challenging. For the same reasons, the framework does not allow the combination of uncertainty information from multiple data sources using different uncertainty approaches. Sentences should be annotated using the same uncertainty approaches for the calculus to be executed, otherwise, this presents no problem with other metadata since they will be appended to the presented information using the *metaList* algorithm. Providing a solution for the latter problem, we can add a third step to the previous two-step process in subsection 4.1 corresponding to the combination of uncertainty information of identical sentences issued from different contexts with different uncertainty information. In any case, the *metaList* algorithm can be adapted to different types of metadata (ex: provenance).

A security problem about altering information can reach uncertainty calculi: altering uncertainty operations may alter the ranking of uncertain linked data and the newly generated data from it.

To the best of our knowledge, there is no previous work that offered a framework to handle uncertainty using a generic standard vocabulary to represent custom uncertainty approaches. Using our approach, uncertainty is represented using the Semantic Web standards to be publishable and reusable. In addition, the use of contexts and calculi allows more selectivity towards the metadata presented to the user and allows inferring new uncertainty information.

mUnc can extend the work by Cabrio et al. [30] by enriching the proposed fuzzy labeling algorithms with definitions of uncertainty theories suitable to the data. The Linked Data sources can adopt this approach to enrich federated

queries with uncertainty information and, progressively, build a consensus-based Linked Data source. A set of other applications such as fake news detection (definition of a theory and logic for fake news), argumentation-based systems and even community-based datasources such as DBpedia can use *mUnc* to enrich their future content with uncertainty information.

6 Conclusion

In this paper, we discussed the representation and publication of uncertainty on the Semantic Web. We presented a vocabulary allowing the representation of uncertainty theories and the annotation of sentences using the Semantic Web standards. We explained the publishing of a reusable uncertainty calculi using *LDScript* and *Corese*.

Uncertainty representation is the first step of a long process, including the preliminary calculus of uncertainty values, the propagation of uncertainty among interconnecting sources. We plan to add the possibility of merging uncertain data annotated using different uncertainty approaches. We would like to implement context overlapping relying on a set of properties such as weights and thresholds, allowing the selectivity between context metadata and the optimization of the storage. Finally, this work makes a step towards inconsistency-tolerant and context-aware uncertain Web of data.

References

1. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Scientific american* **284**(5), 34–43 (2001)
2. Bizer, C., Heath, T., Berners-Lee, T.: Linked data: The story so far. In: *Semantic services, interoperability and web applications: emerging concepts*, pp. 205–227. IGI Global (2011)
3. Bollacker, K., Evans, C., Paritosh, P., Sturge, T., Taylor, J.: Freebase: a collaboratively created graph database for structuring human knowledge. In: *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. pp. 1247–1250. AcM (2008)
4. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering* (6), 734–749 (2005)
5. Mohirta, M., Cernian, A., Carstoiu, D., Vladu, A.M., Olteanu, A., Sgarciu, V.: A semantic Web based scientific news aggregator. In: *Applied Computational Intelligence and Informatics (SACI), 2011 6th IEEE International Symposium on*. pp. 285–289. IEEE (2011)
6. Gandon, F., Governatori, G., Villata, S.: Normative Requirements as Linked Data. In: *The 30th international conference on Legal Knowledge and Information Systems (JURIX 2017)* (2017)
7. Benslimane, D., Sheng, Q.Z., Barhamgi, M., Prade, H.: The uncertain Web: Concepts, challenges, and current solutions. *ACM Transactions on Internet Technology (TOIT)* **16**(1), 1 (2016)

8. Dubois, D., Prade, H.: Formal representations of uncertainty. *Decision-Making Process: Concepts and Methods* pp. 85–156 (2009)
9. Klir, G.J., Smith, R.M.: On measuring uncertainty and uncertainty-based information: recent developments. *Annals of Mathematics and Artificial Intelligence* **32**(1-4), 5–33 (2001)
10. Laskey, K.J., Laskey, K.B.: Uncertainty Reasoning for the World Wide Web: Report on the URW3-XG Incubator Group. In: URSW. Citeseer (2008)
11. Reynolds, D.: Uncertainty reasoning for linked data. In: *Proceedings of the Fifth International Conference on Uncertainty Reasoning for the Semantic Web-Volume 527*. pp. 85–88. CEUR-WS. org (2009)
12. Hayes, P.: RDF semantics, W3C recommendation. <http://www.w3.org/TR/rdf-mt/> (2004)
13. McGlothlin, J.P., Khan, L.R.: Materializing and Persisting Inferred and Uncertain Knowledge in RDF Datasets. In: *AAAI*. vol. 10, pp. 11–15 (2010)
14. d’Amato, C., Bryl, V., Serafini, L.: Semantic Knowledge Discovery and Data-Driven Logical Reasoning from Heterogeneous Data Sources. In: *Uncertainty Reasoning for the Semantic Web III*, pp. 163–183. Springer (2014)
15. Ding, Z., Peng, Y., Pan, R.: BayesOWL: Uncertainty modeling in semantic web ontologies. In: *Soft computing in ontologies and semantic web*, pp. 3–29. Springer (2006)
16. Stoilos, G., Stamou, G.B., Tzouvaras, V., Pan, J.Z., Horrocks, I.: Fuzzy OWL: Uncertainty and the Semantic Web. In: *OWLED* (2005)
17. Safia, B.B., Aicha, M.: Poss-OWL 2: Possibilistic Extension of OWL 2 for an uncertain geographic ontology. *Procedia Computer Science* **35**, 407–416 (2014)
18. Dividino, R., Sizov, S., Staab, S., Schuele, B.: Querying for provenance, trust, uncertainty and other meta knowledge in RDF. *Web Semantics: Science, Services and Agents on the World Wide Web* **7**(3), 204–219 (2009)
19. Bouquet, P., Serafini, L., Stoermer, H.: Introducing Context into RDF Knowledge Bases. In: *SWAP*. vol. 5, pp. 14–16 (2005)
20. Carroll, J.J., Bizer, C., Hayes, P., Stickler, P.: Named graphs, provenance and trust. In: *Proceedings of the 14th international conference on World Wide Web*. pp. 613–622. ACM (2005)
21. Consortium, W.W.W., et al.: RDF 1.1 concepts and abstract syntax (2014)
22. Brickley, D., Guha, R.V., McBride, B.: RDF Schema 1.1. W3C recommendation **25**, 2004–2014 (2014)
23. Corby, O., Faron-Zucker, C.: RDF/SPARQL Design Pattern for Contextual Metadata. In: *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*. pp. 470–473. WI ’07, IEEE Computer Society, Washington, DC, USA (2007)
24. Corby, O., Faron-Zucker, C., Gandon, F.: LDScript: a Linked Data Script Language. In: *International Semantic Web Conference*. pp. 208–224. Springer (2017)
25. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: DBpedia-A crystallization point for the Web of Data. *Web Semantics: science, services and agents on the world wide web* **7**(3), 154–165 (2009)
26. Cabrio, E., Villata, S., Gandon, F.: Classifying Inconsistencies in DBpedia Language Specific Chapters. In: *LREC*. pp. 1443–1450 (2014)
27. Klir, G., Yuan, B.: *Fuzzy sets and fuzzy logic*, vol. 4. Prentice hall New Jersey (1995)
28. da Costa Pereira, C., Dubois, D., Prade, H., Tettamanzi, A.G.: Handling Topical Metadata Regarding the Validity and Completeness of Multiple-Source Informa-

- tion: A Possibilistic Approach. In: International Conference on Scalable Uncertainty Management. pp. 363–376. Springer (2017)
29. Corby, O., Zucker, C.F.: Corese: A corporate semantic web engine. In: International Workshop on Real World RDF and Semantic Web Applications, International World Wide Web Conference (2002)
30. Cabrio, E., Villata, S., Palmero Aprosio, A.: A RADAR for information reconciliation in Question Answering systems over Linked Data 1. *Semantic Web* **8**(4), 601–617 (2017)